Expectation of a Random Variable

- We define the expectation of a (discrete) random variable X as $\sum_x x \cdot \Pr(X = x)$

Example: Expected Waiting Time for a "Heads"

- Let the random variable X denote the number of flips of a p-biased coin until we get the first "heads"
  - $\Pr(X = i) = p(1-p)^{i-1}$ for $i \geq 1$
  - Thus $\mathrm{E}(X) = \sum_{i \geq 1} i \cdot p(1-p)^{i-1}$
- For any z in $[0, 1)$, we have $S = \sum_{i \geq 1} i \cdot z^i = \frac{z}{(1-z)^2}$
  - Observe that $S - zS = \sum_{i \geq 1} z^i = \frac{z}{1-z}$
- Thus $\mathrm{E}(X) = \frac{p}{1-p} \cdot \frac{1-p}{p^2} = \frac{1}{p}$

Linearity of Expectation

- For any random variables X and Y, we have

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$$

$$\begin{aligned}
\mathrm{E}(X+Y) &= \sum_{x,y} (x+y) \cdot \Pr(X = x \wedge Y = y) \\
&= \sum_{x,y} x \cdot \Pr(X = x \wedge Y = y) + \sum_{x,y} y \cdot \Pr(X = x \wedge Y = y) \\
&= \sum_{x} x \cdot \Pr(X = x) + \sum_{y} y \cdot \Pr(Y = y) \\
&= \mathrm{E}(X) + \mathrm{E}(Y)
\end{aligned}$$

Expected Number of Flips to get Heads (Revisited)

- Let the "indicator" random variable X_i be equal to 1 if the number of flips is at least i, and 0 otherwise
  - The X_i's are not independent
  - We have $\Pr(X_i = 1) = (1-p)^{i-1}$ and thus $\mathrm{E}(X_i) = (1-p)^{i-1}$
- Observe that $X = \sum_{i \geq 1} X_i$
- Thus, by linearity of expectation, we have

$$\mathrm{E}(X) = \sum_{i \geq 1} (1-p)^{i-1} = \frac{1}{1-(1-p)} = \frac{1}{p}$$

Example: A Coupon Collector Process

- Suppose we repeatedly draw a uniformly random number from $\{1, \ldots, n\}$ until we have drawn each number at least once
- What is the expected number of draws?
- We partition the analysis into n phases
  - For $1 \leq i \leq n$, phase i begins (and phase $i-1$ terminates) once $i-1$ distinct integers have been drawn
  - Let the random variable X_i denote the number of draws in phase i
  - Let X denote $\sum_{1 \leq i \leq n} X_i$

A Coupon Collector Process (cont'd)

- The random variable X_i corresponds to the number of flips to get heads with a $\frac{n-i+1}{n}$-biased coin
- Thus $\mathrm{E}(X_i) = \frac{n}{n-i+1}$
- By linearity of expectation, we have

$$\mathrm{E}(X) = \sum_{1 \le i \le n} \frac{n}{n-i+1} = n \sum_{1 \le i \le n} \frac{1}{i} = nH_n \sim n \ln n$$

Randomized Quicksort

- Consider a randomized variant of Quicksort in which the pivot is chosen uniformly at random
- Assume for the sake of convenience that the n keys to be sorted are distinct
- Index the keys from 1 to n from smallest to largest
- Let the random variable X denote the number of comparisons performed
- For $1 \leq i < j \leq n$, let the indicator random variable $X_{i,j}$ be equal to 1 if keys i and j are compared, and 0 otherwise
- Since no pair of keys are compared more than once, we have $X = \sum_{1 \leq i < j \leq n} X_{i,j}$

Randomized Quicksort (cont'd)

- For any i and j such that $1 \leq i < j \leq n$, what is the probability that keys i and j are compared?
  - They are compared if and only if the first pivot drawn from the set of keys with indices in $\{i, \ldots, j\}$ is either i or j
  - By symmetry, this occurs with probability $\frac{2}{j-i+1}$
- By linearity of expectation, we have

$$E(X) = \sum_{1 \leq i < j \leq n} \frac{2}{j - i + 1}$$

Randomized Quicksort (cont'd)

- We need to evaluate $\sum_{1 \leq i < j \leq n} \frac{1}{j-i+1}$
  - This sum is $(n-1) \cdot \frac{1}{2} + (n-2) \cdot \frac{1}{3} + \ldots + 1 \cdot \frac{1}{n}$
- Thus

$$
\begin{aligned}
\mathrm{E}(x) &= 2 \sum_{2 \leq i \leq n} \frac{n-i+1}{i} \\
&= 2(n+1)(H_n - 1) - 2(n-1) \\
&= 2(n+1)H_n - 2n - 2 - 2n + 2 \\
&= 2(n+1)H_n - 4n \\
&\sim 2n \ln n
\end{aligned}
$$

Markov's Inequality

- For any nonnegative random variable X and any $a > 0$, we have
$$\Pr(X \geq a) \leq \frac{\mathrm{E}(X)}{a}$$

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{x} x \cdot \Pr(X = x) \\
&\geq \left(\sum_{x<a} 0 \cdot \Pr(X = x)\right) + \left(\sum_{x \geq a} a \cdot \Pr(X = x)\right) \\
&= a \sum_{x \geq a} \Pr(X = x) \\
&= a \cdot \Pr(X \geq a)
\end{aligned}$$

Expectation of a Product of Independent RVs

- If random variables X and Y are independent, then $\mathrm{E}(XY) = \mathrm{E}(X) \cdot \mathrm{E}(Y)$

$$
\begin{aligned}
\mathrm{E}(XY) &= \sum_{x,y} xy \cdot \Pr(X = x \wedge Y = y) \\
&= \sum_{x,y} xy \cdot \Pr(X = x) \Pr(Y = y) \\
&= \sum_{x} x \cdot \Pr(X = x) \sum_{y} y \cdot \Pr(Y = y) \\
&= \mathrm{E}(Y) \sum_{x} x \cdot \Pr(X = x) \\
&= \mathrm{E}(X) \cdot \mathrm{E}(Y)
\end{aligned}
$$

Variance of a Random Variable

- The variance of a random variable X, denoted $\text{Var}(X)$, is defined as the expected value of the square of the difference between X and its expectation

$$
\begin{aligned}
\text{Var}(X) &= \text{E}[(X - \text{E}(X))^2] \\
&= \text{E}[X^2 - 2X\text{E}(X) + \text{E}(X)^2] \\
&= \text{E}(X^2) - 2\text{E}(X)\text{E}(X) + [\text{E}(X)]^2 \\
&= \text{E}(X^2) - [\text{E}(X)]^2
\end{aligned}
$$

Example: The Variance of a p-Biased Coin Flip

- Let the random variable X be equal to 1 with probability p, and to 0 with probability $1 - p$
- Thus X^2 has the same distribution as X since $1^2 = 1$ and $0^2 = 0$
- Hence

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 = p - p^2 = p(1 - p)$$

Variance of a Sum of Independent Random Variables

- If random variables X and Y are independent, then $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathrm{E}[(X+Y)^2] - [\mathrm{E}(X+Y)]^2 \\
&= \mathrm{E}(X^2) + 2\mathrm{E}(XY) + \mathrm{E}(Y^2) - [\mathrm{E}(X) + \mathrm{E}(Y)]^2 \\
&= \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 + \mathrm{E}(Y^2) - [\mathrm{E}(Y)]^2 \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y)
\end{aligned}$$

Example: The Variance of a Series of Coin Flips

- Let X denote the number of heads in n independent tosses of a p-biased coin
- Let the random variable X_i be equal to 1 if the ith toss comes up heads, and 0 otherwise
- Since the X_i's are independent, we have

$$\mathrm{Var}(X) = \sum_{1 \leq i \leq n} \mathrm{Var}(X_i) = np(1 - p)$$

Chebyshev's Inequality

- For any random variable X and any $a > 0$, we have

$$\Pr(|X - \mathrm{E}(X)| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}$$

  - The LHS is equal to $\Pr(Y \geq a^2)$ where Y denotes the nonnegative random variable $(X - \mathrm{E}(X))^2$
  - Thus Markov's inequality implies that the LHS is at most $\mathrm{E}(Y)/a^2$
  - The claim follows since $\mathrm{E}(Y) = \mathrm{Var}(X)$

The Binomial Distribution

- In the design and analysis of randomized algorithms, we often encounter binomially distributed random variables
  - Counts the number of successes in n independent trials with success probability p
  - We let $B(n, p)$ denote this distribution
- We can give derive sharp tail bounds for this specific distribution
  - These tail bounds also apply to the case where the n independent trials have average success probability p

Tail Bounds for the Binomial Distribution

- For $1 \leq i \leq n$, let p_i belong to $[0, 1]$, and let X_i be a 0-1 random variable such that $\Pr(X_i = 1) = p_i$
- Let p denote $\frac{1}{n}\sum_{1 \leq i \leq n} p_i$ and let the random variable X denote $\sum_{1 \leq i \leq n} X_i$
- Thus $\mathrm{E}(X) = np$
- In what follows, we develop several useful "Chernoff bounds" on the upper and lower tail of X

A Useful Inequality

- Lemma 1: $E[\exp(tX)] \leq \exp[np(e^t - 1)]$ for all t
  - Using $1 + x \leq e^x$ for all x, we have

$$\begin{aligned}
E(\exp(tX)) &= E\left[\prod_{1 \leq i \leq n} \exp(tX_i)\right] \\
&= \prod_{1 \leq i \leq n} E[\exp(tX_i)] \\
&= \prod_{1 \leq i \leq n} [p_i e^t + (1 - p_i)] \\
&= \prod_{1 \leq i \leq n} [1 + p_i(e^t - 1)] \\
&\leq \prod_{1 \leq i \leq n} \exp[p_i(e^t - 1)] \\
&= \exp[np(e^t - 1)]
\end{aligned}$$

A Bound on the Lower Tail

- For any δ in $[0, 1)$ and any $t \geq 0$, Markov's inequality implies

$$\begin{aligned}
\Pr(X \leq (1-\delta)np) &= \Pr(-X \geq -(1-\delta)np) \\
&= \Pr(\exp(-tX) \geq exp(-t(1-\delta)np)) \\
&\leq \frac{\mathrm{E}[\exp(-tX)]}{exp(-t(1-\delta)np)}
\end{aligned}$$

- Applying Lemma 1 with $t = -\ln(1-\delta) \geq 0$, we obtain

$$\Pr(X \leq (1-\delta)np) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{np}$$

A Bound on the Lower Tail (cont'd)

- For δ in $[0, 1)$, we have

$$\ln(1-\delta) = -\delta - \frac{\delta^2}{2} - \frac{\delta^3}{3} - \dots$$

- Thus

$$\begin{aligned}(1-\delta)\ln(1-\delta) &= -\delta + \left(1 - \frac{1}{2}\right)\delta^2 + \left(\frac{1}{2} - \frac{1}{3}\right)\delta^3 + \dots \\ &\geq -\delta + \delta^2/2\end{aligned}$$

- Hence

$$(1-\delta)^{1-\delta} = \exp[(1-\delta)\ln(1-\delta)] \geq \exp(-\delta + \delta^2/2)$$

A Bound on the Lower Tail (cont'd)

- Combining the preceding inequalities, we obtain the following convenient bound for δ in $[0, 1)$

$$
\begin{aligned}
\Pr(X \leq (1-\delta)np) &\leq \left(\frac{e^{-\delta}}{\exp(-\delta + \delta^2/2)} \right)^{np} \\
&= \exp(-\delta^2 np/2)
\end{aligned}
$$

A Bound on the Upper Tail

- For any nonnegative δ and t, Markov's inequality implies

$$\begin{aligned}
\Pr(X \geq (1+\delta)np) &= \Pr(\exp(tX) \geq \exp(1+\delta)npt) \\
&\leq \frac{\mathrm{E}(\exp(tX))}{\exp((1+\delta)npt)}
\end{aligned}$$

- Using Lemma 1 with $t = \ln(1+\delta) \geq 0$, we obtain

$$\Pr(X \geq (1+\delta)np) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{np}$$

A Bound on the Upper Tail for "Small Deviations"

- The bound we have derived on the upper tail is a bit messy
- Proceeding in a similar manner as in the case of the lower tail bound, we can derive the following simpler bound for δ in $[0, 1)$

$$\Pr(X \geq (1+\delta)np) \quad \leq \quad \exp(-\delta^2 np/3)$$

- Note that the "2" appearing in the exponent of the lower tail bound derived earlier is weakened to a "3" in the above bound

A Stronger Bound for an Important Special Case

- Oftentimes we are interested in the case of n fair coin flips, i.e., $B(n, 1/2)$
- For the special case where X is drawn from $B(n, 1/2)$, we can use a similar approach to derive the following bound for all δ in $[0, 1]$
$$\Pr(X \geq (1+\delta)n/2) \leq \exp(-\delta^2 n/2)$$
- By symmetry, we also have
$$\Pr(X \leq (1-\delta)n/2) \leq \exp(-\delta^2 n/2)$$
for all δ in $[0, 1]$

Application: Tossing a Fair Coin

- Let X be drawn from $B(n, 1/2)$
- Suppose we wish to upper bound $q = \Pr(X \geq (n/2) + c\sqrt{n})$
- Markov's inequality gives

$$q \leq \frac{n/2}{(n/2) + c\sqrt{n}} = \frac{1}{1 + \frac{2c}{\sqrt{n}}} \sim 1 - \frac{2c}{\sqrt{n}} = 1 - o(1)$$

- Chebyshev's inequality gives

$$q \leq \Pr(|X - n/2| \geq c\sqrt{n}) \leq \frac{n/4}{c^2 n} = \frac{1}{4c^2}$$

- We can apply the Chernoff bound for $B(n, 1/2)$ with $\delta = \frac{2c}{\sqrt{n}}$ to obtain

$$q \leq \exp(-2c^2)$$

Some "Balls and Bins" Problems

- Suppose we throw a series of balls independently and uniformly at random into n bins
- The coupon collector problem considered earlier tells us the expected number of balls required to "collect" each bin
- How many balls do we need to throw to ensure that with high probability each bin receives at least one ball?
- If we throw n balls, what bounds can we prove on the maximum load of any bin?
- How many balls do we need to throw to ensure that ratio between the maximum and minimum bin loads is at most $1 + \varepsilon$?

Upper Bounding the Load of a Specific Bin

- Assume that n balls are thrown independently and uniformly at random into n bins
- Let the random variable X denote the load of bin 1 (say)
- Let X_i denote the indicator random variable that is equal to 1 if ball i lands in bin 1, and to 0 otherwise
  - Thus $X = \sum_{1 \leq i \leq n} X_i$
  - Thus X is drawn from $B(n, 1/n)$, and has expectation 1
  - Using the large deviation Chernoff bound, we can show that for any positive constant c, there is a positive constant c' such that
$$\Pr\left(X \geq c' \frac{\ln n}{\ln \ln n}\right) \leq n^{-c}$$

Upper Bounding the Maximum Bin Load

- Let E_i denote the event that the load of bin i exceeds $c'f(n)$ where $f(n) = \frac{\ln n}{\ln \ln n}$
- We have seen that $\Pr(E_1) \leq n^{-c}$
- By symmetry, $\Pr(E_i) \leq n^{-c}$ for any bin i
- By a union bound, the probability that any of the "bad events" E_i occurs is n^{1-c}
- We conclude that the maximum load of any bin is $O(f(n))$ with high probability
- Is this bound tight?

Lower Bounding the Load of a Specific Bin

- Since $\binom{n}{k} = \frac{n}{k} \cdots \frac{n-k+1}{1} \geq (n/k)^k$, we have

$$\begin{aligned}
\Pr(X = k) &= \binom{n}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{n-k} \\
&\geq k^{-k} \left(1 - \frac{1}{n}\right)^{n-k} \\
&= \Omega(k^{-k})
\end{aligned}$$

- Using the above, we can show that for any positive constant ε, there is a positive constant ε' such that $\Pr(X \geq \varepsilon' f(n)) \geq n^{-\varepsilon}$

Lower Bounding the Minimum Load

- For any bin i, let E_i denote the event that bin i receives at least $\varepsilon' f(n)$ balls
  - We have shown that $\Pr(E_1) \geq n^{-\varepsilon}$
  - By symmetry, we have $\Pr(E_i) \geq n^{-\varepsilon}$ for any bin i
- If the events E_i were independent, we could upper bound the probability that none of the E_i's occur by $(1 - n^{-\varepsilon})^n$ which is approximately $\exp(-n^{1-\varepsilon})$, and hence is super-polynomially small
- Unfortunately, the events E_i are correlated

Exploiting Negative Correlation of the E_i's

- Observe that the following inequality holds for any bin i

$$\Pr(E_i \mid \cap_{1 \le j < i} \neg E_j) \ge \Pr(E_i)$$

- Thus

$$\begin{aligned}\Pr(\cap_{1 \le i \le n} \neg E_i) &= \prod_{1 \le i \le n} \Pr(\neg E_i \mid \cap_{1 \le j < i} \neg E_j) \\ &\le \prod_{1 \le i \le n} \Pr(\neg E_i)\end{aligned}$$

- In other words, we can get a valid upper bound on $\Pr(\cap_{1 \le i \le n} \neg E_i)$ by treating the events E_i as if they were independent